

# Method of autoregression in application of singular-spectral analysis of time series for forecasting production of oil and gas industry products

*Aleksandr Zhukov*<sup>1,2</sup>, *Denis Zhurenkov*<sup>3</sup>, *Yuliya Basalaeva*<sup>3</sup>, *Victor Gedzyun*<sup>1</sup>,  
*Igor Kartsan*<sup>4,5\*</sup>, and *Iuliia Dementeva*<sup>6</sup>

<sup>1</sup>The Federal Center of Expertize and Analyzis, 33, p. 4, Talalikhina str., Moscow, 109316, Russia

<sup>2</sup>Institute of Astronomy of the Russian Academy of Sciences, 48, Pyatnitskaya str., Moscow, 119017, Russia

<sup>3</sup>Federal state unitary enterprise «All-Russia scientific and research institute «Center», 11, p.1., Sadovaya-Kudrinskaya str., Moscow, 123242, Russia

<sup>4</sup>Marine Hydrophysical Institute, Russian Academy of Sciences, 2, Kapitanskaya str., Sevastopol, 299011, Russia

<sup>5</sup>Reshetnev Siberian State University of Science and Technology, 31, Krasnoïarskii Rabochii Prospekt, Krasnoyarsk, 660037, Russia

<sup>6</sup>Lavochkin Association, 24, Leningradskaya str., Khimki, Moscow region, 141402, Russia

**Abstract.** More and more time series data are produced in various fields. It provides data for the research of time series analysis method, and promotes the development of time series research. Due to the generation of highly complex and large-scale time series data, the construction of forecasting models for time series data brings greater challenges. The theoretical aspects of using the model of singular-spectral analysis of time series with the use of autoregression are considered, and the justification of the expediency of using this model for forecasting the production of products for both the oil and gas industry and dual-use products is given. Both autoregressive model and decision tree model can be applied with the same degree of reliability for forecasting aggregate values of production.

## 1 Introduction

The basis of any automated time series forecasting is the construction of a model to calculate future events based on known events of the past.

From the existing models of time series in the theory of econometrics two models were chosen, as they best meet the requirements for the task of forecasting socio-economic macro-indicators in the medium and long term forecasting of industries:

- A model based on the transformation of non-stationary series into time components using the singular value decomposition method with the application of decision trees.

---

\* Corresponding author: [kartsan2003@mail.ru](mailto:kartsan2003@mail.ru)

- Model based on the transformation of non-stationary series into time components using the singular expansion method with the application of autoregressive model.

The mechanism of action of singular-spectral analysis of time series for forecasting the production of dual-purpose products, as well as the training of the "noise" component through the application of decision trees is most fully described in [1-5].

In machine learning methods there are a large number of different models and approaches for their formation, but all are united by the need to calculate the initial parameters of the model. The computation of these parameters requires large time expenditures, so their reduction is one of the most important problems, for which there are various methods. One of such approaches is the use of parallel algorithms [6]. In this article we will talk about the mechanisms of action and analyze the application of the autoregressive model, or, more precisely, its variant, the autoregressive integrated moving average (ARIMA) model, which is based on the parallel execution of two independent processes.

## 2 Theoretical aspects of using autoregressive models

Autoregressive model is a method of modeling the subsequent components based on the previous ones in the form of a linear function. Autoregressive model in its pure form is rarely used in practice, as it is suitable only for time series without trend and seasonal components.

The most advanced variants are such models as: autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), seasonal autoregressive integrated moving averages (SARIMA), seasonal autoregressive integrated moving averages with exogenous regressors (SARIMAX), vector autoregression (VAR), moving average of vector autoregression (VARMA), moving average of vector autoregression with exogenous regressors (VARMAX).

This study will analyze the application of ARIMA model (Box-Jenkins model).

When analyzing historical data, there are no distinct regular components in the time series of dual-use products. In addition, individual observations may contain significant errors, which interferes with the identification of regular components, as well as with the construction of the forecast. Autoregressive integrated moving average (ARIMA) allows the selection of regular components [7-15].

The ARIMA model is a model for analyzing non-stationary time series that combines two parallel processes - autoregression and moving average. Based on the definition, the mechanism of the ARIMA model can be divided into two processes.

Autoregressive (AR) process. Provided that there are elements in the time series that consistently depend on each other, the time series can be written as a function:

$$y_t = z + a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + \dots + \varepsilon \quad (1)$$

where  $z$  is a constant,  $a_1, a_2, a_3$  are autoregressive parameters.

Each element of the series is the sum of a random component as well as a linear combination of previous observations [16-19].

Moving average (MA) process. The difference between the moving average process and the autoregressive process is the presence of the cumulative effect of previous errors on each subsequent observation. MA takes the following form:

$$y_t = \mu + e_t - b_1 \varepsilon_{t-1} - b_2 \varepsilon_{t-2} + b_3 \varepsilon_{t-3} \quad (2)$$

where  $\mu$  is a constant,  $b_1, b_2, b_3$  are autoregressive parameters. That is, the series model is a sum of the random component at the current moment of time and a linear combination of random variables at previous moments of time.

As a rule, ARIMA models (p, d, q) are used, whose parameter values do not exceed 2. In this case, the parameter p determines the order of the autoregressive component, the parameter q determines the order of the moving average, and the parameter d determines the order of the difference (discrete derivative).

Thus, the ARIMA model is a model combining two processes (AR and MA) and has the following form:

$$\Delta^d_{y_t} = c + \sum_{i=1}^p a_i \mu * \Delta^d_{y_{t-1}} + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t \quad (3)$$

where  $\varepsilon_t$  is the stationary time series,  $c, a_i, b_j$  are the model parameters,  $\Delta^d$  is the difference operator.

The main advantages of using ARIMA models are:

- the greatest scientific validity and clear statistical-mathematical justification compared to other time series forecasting models;
- simplicity in application to time series forecasting, availability of formalized instructions for selection of the most appropriate model.

Among the disadvantages of the ARIMA model we can single out the presence of strict requirements for the number of observations, non-adaptability and the need for re-estimation when data change, as well as large time costs.

### 3 Model comparison

To analyze the adequacy of the application of the model of singular-spectral analysis of time series with the use of autoregression and with the use of decision trees to forecasting the production of products, including dual-use, the forecast of organizations by industry was compared with the forecast of artificial intelligence according to the two models and the corresponding deviations were identified.

To conduct the study, a sample of 46 nomenclature items of products manufactured by organizations, which most fully reflect the general picture of production of products, including dual-use defense industries, was made, of which:

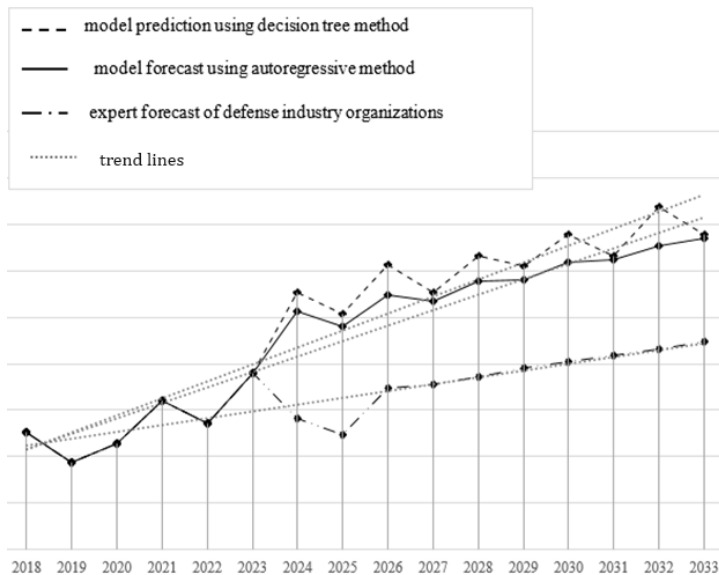
- 15 nomenclature items produced by organizations of the aviation industry, including: navigation, meteorological and geophysical equipment, systems and complexes of on-board radio electronic equipment (hereinafter - BREE), autopilot equipment, fire alarm systems and other high-tech products;
- 21 nomenclature items produced by organizations of the radio electronic industry, including: radio measuring equipment, microwave electronics and related components, components of electrical machines, vacuum devices, complexes of devices for pre-flight inspection of avionics and other high-tech products;
- 10 nomenclature items produced by oil and gas industry organizations, including: marine instrumentation products (engines, communication equipment, electromagnets, electromagnetic contactors and starters, navigation, meteorological, geophysical instruments and similar tools), related products (lighting fixtures) and other high-tech products.

The input data for the forecast were the actual data on the output of products including dual-use products for 2018-2022, as well as the estimated value for 2023. The forecast values were built for the period up to 2033. To summarize the situation, we present a comparison of aggregated data by industry.

Aviation industry. Based on the comparison of forecast data for 17 key products of the aviation industry to which the artificial intelligence forecasting mechanism was applied, the following conclusions can be drawn (Figure 1):

- linear trends of forecasts for the period up to 2033 have the same directionality and do not diverge in the long-term period;
- in the short-term period, the model forecasts may significantly deviate from the expert forecasts of the defense industry organizations due to the fact that the model does not take into account the existing plans of the defense industry organization, as well as political and economic risks;

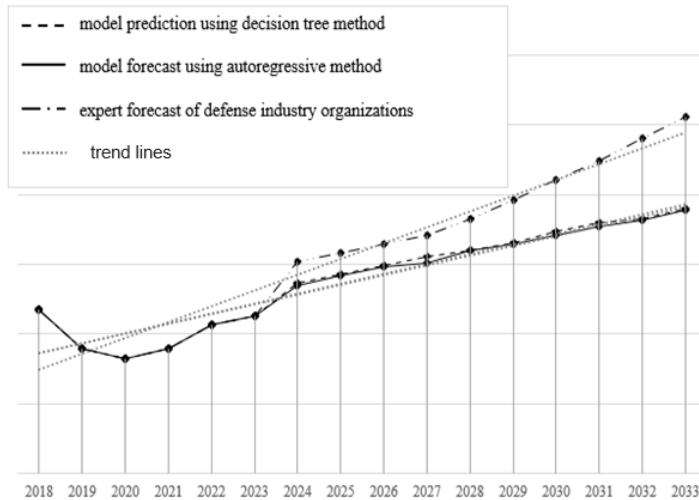
- the expert forecast of defense industry organizations is more conservative compared to forecast modeling, which may be due to intentional underestimation of values in order to achieve (exceed) indicators in the future;
- on average, the forecast using decision trees deviates by 67%, and the forecast using autoregression - by 57% to the greater side from the expert forecast of the defense industry organizations;
- the average growth rate for the period 2018-2033 according to the expert forecast of DIC organizations was 104.2%, according to the forecast modeling using decision trees - 107.3%, according to the forecast modeling using autoregression - 107.2%.



**Fig. 1.** Comparison of aggregated forecasts for dual-use products of aviation industry organizations.

Radioelectronic industry. Based on the comparison of forecast data for 21 key products of the radio electronics industry, to which the forecasting mechanism with the help of artificial intelligence was applied, the following conclusions can be drawn (Figure 2):

- linear trends of forecasts in the short-term and long-term period have the same direction and do not diverge;
- on average, the forecast using decision trees deviates by 16%, and the forecast using autoregressive model - by 17% downward from the expert forecasts of defense industry organizations;
- the average growth rate for the period 2018-2033 according to expert forecasts of ICS organizations was 105.7%, while according to forecast modeling using decision trees - 103.5%, and by forecast modeling using autoregression - 103.4%.



**Fig. 2.** Comparison of aggregated forecasts for dual-use products of radioelectronic industry organizations.

Oil and gas industry. Based on the comparison of forecast data for 10 key products of the oil and gas industry to which the artificial intelligence forecasting mechanism was applied, the following conclusions can be drawn (Figure 3):

- in general, the linear trends of forecasts have the same directionality and do not diverge over the long-term period;
- in the short-term period the model forecasts may significantly deviate from the expert forecasts of the defense industry organizations, which is due to the cyclical nature of shipbuilding production;
- expert forecasts of defense industry organizations are more optimistic compared to forecast modeling;
- on average, for the entire period, the forecast using decision trees deviates by 6.6%, and the forecast using autoregression - by 6.3% to the lesser side from the expert forecasts of the defense industry organizations;
- the average growth rate for the period 2018-2033 according to the expert forecast of the DIC organizations was 123.8%, according to the forecast modeling using decision trees - 119.8%, and using autoregression - 120.1%.

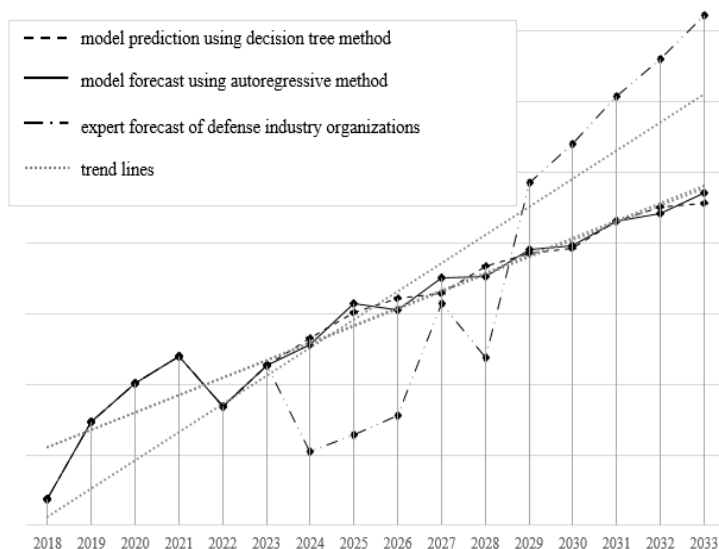
In addition, the RMSE (root mean square error) metric was calculated to select the most appropriate research model. RMSE is calculated using the standard formula:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{4}$$

As a rule, in order to better understand the values of the RMSE indicator, it is normalized by the following formula:

$$RMSE_{\text{normalized}} = \frac{RMSE}{\text{max row value} - \text{min row value}} \tag{5}$$

Normalization brings the entire range of values to between 0 and 1, which allows models with closer to zero RMSE values to be interpreted as better fit.



**Fig. 3.** Comparison of aggregate forecasts for products of oil and gas industry organizations.

## 4 Conclusion

The model of singular-spectral analysis of time series with application of decision trees and the model of singular-spectral analysis of time series with application of autoregression can be used for forecasting the production of dual-use products by the organizations of the defense industry, as they show sufficiently accurate and adequate results in the long-term period and do not contradict the vision of the development of product output by experts of the defense industry organizations.

The work was carried out within the framework of the state assignment of the Ministry of Education and Science of Russia on the topic "Development of methodology of production of dual-use products by high-tech companies of Russia using elements of artificial intelligence in the conditions of digitalization of the economy and sanctions pressure" № 123011600034-3.

## References

1. E.V. Malovetskaya, A.K. Mozalevskaya, T-Comm: Telecommunications and Transport **17(1)**, 33-41 (2023)
2. I.A. Merkulov, In collection: Information processing and mathematical modeling. materials of the All-Russian scientific and technical conference with international participation (Novosibirsk), pp. 179-184 (2023)
3. N.V. Abalov, V.V. Gubarev, O.K. Alsova, Proceedings of SPIIRAN **4(35)**, 49-63 (2014)
4. S.A. Tolkachev, Economic Revival of Russia **3(61)**, 127-145 (2019)
5. I. Kartsan, A. Zhukov, S. Pronichkin, E3S Web of Conferences **376**, 05013 (2023). <https://doi.org/10.1051/e3sconf/202337605013>
6. A. Zhukov, A. Zaverzaev, G. Sozinov, I. Kartsan, AIP Conference Proceedings **246722**, 020025 (2021). <https://www.doi.org/10.1063/5.0092533>
7. A.I. Stankevich, Volga region. Physico-mathematical sciences **1(9)**, 134-138 (2009)

8. A.M. Batkovskiy, M.A. Batkovskiy, P.V. Kravchuk, Actual issues of modern economics **3**, 267-279 (2020)
9. V.P. Bauer, S.N. Silvestrov, P.Yu. Baryshnikov, Information Society **3**, 30-40 (2017)
10. S.I. Dovguchits, D.A. Zhurenkov, Scientific Bulletin of the defense-industrial complex of Russia **3**, 16-24 (2022). [https://www.doi.org/10.52135/2410-4124\\_2022\\_3\\_16](https://www.doi.org/10.52135/2410-4124_2022_3_16)
11. S.I. Dovguchits, D.A. Zhurenkov, Scientific Bulletin of the defense-industrial complex of Russia **4**, 5-12 (2022)
12. A.A. Grishin, S.P. Stroev, Vestnik of Technological University **9**, 180-186 (2018)
13. I. Kartsan, A. Zhukov, S. Pronichkin, E3S Web of Conferences **389**, 090056 (2023). <https://doi.org/10.1051/e3sconf/202338909056>
14. T.B. Timofeeva, E.A. Ozdoeva, Management **8(3)**, 112-122 (2020). <https://www.doi.org/10.26425/2309-3633-2020-8-3-112-122>
15. S.I. Dovguchits, Scientific Bulletin of the Russian Defense Industry Complex **4**, 39-42 (2020)
16. N. Voenkova, Ekonomicheskie strategii **8**, 74-77 (2015)
17. I. Trauschweizer, Journal of Military History **80(1)**, 296-298 (2016)
18. Ch.J. Dunlap, Daedalus **140(3)**, 135-147 (2011)
19. O.A. Lukmanov, CADmaster **1(87)**, 55-56 (2018)